Towards unravelling the relationship between on-body, environmental and emotion data using sensor information fusion approach

KANJO, Eiman, YOUNIS, Eman M.G. and SHERKAT, Nasser
<http://orcid.org/0000-0003-1488-5682>

Published version

KANJO, Eiman, YOUNIS, Eman M.G. and SHERKAT, Nasser (2017). Towards unravelling the relationship between on-body, environmental and emotion data using sensor information fusion approach. Information Fusion, 40, 18-31.

Repository use policy

Information Fusion 40 (2018) 18–31

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Full Length Article

Towards unravelling the relationship between on-body, environmental and emotion data using sensor information fusion approach

Eiman Kanjo[a,*], Eman M.G. Younis[b], Nasser Sherkat[a]

[a] Computing and Technology, Nottingham Trent University, United Kingdom
[b] Faculty of Computers and Information Minia University, Egypt

A R T I C L E I N F O

Article history:
Received 18 January 2017
Revised 13 April 2017
Accepted 28 May 2017
Available online 29 May 2017

Keywords:
Multi sensor data fusion
Regression analysis sensor data
Multivariable regression
Affective computing
Physiological signals
Machine learning

A B S T R A C T

Over the past few years, there has been a noticeable advancement in environmental models and information fusion systems taking advantage of the recent developments in sensor and mobile technologies. However, little attention has been paid so far to quantifying the relationship between environment changes and their impact on our bodies in real-life settings.

In this paper, we identify a data driven approach based on direct and continuous sensor data to assess the impact of the surrounding environment and physiological changes and emotion.

We aim at investigating the potential of fusing on-body physiological signals, environmental sensory data and on-line self-report emotion measures in order to achieve the following objectives: (1) model the short term impact of the ambient environment on human body, (2) predict emotions based on-body sensors and environmental data.

To achieve this, we have conducted a real-world study ‘in the wild’ with on-body and mobile sensors. Data was collected from participants walking around Nottingham city centre, in order to develop analytical and predictive models.

Multiple regression, after allowing for possible confounders, showed a noticeable correlation between noise exposure and heart rate. Similarly, UV and environmental noise have been shown to have a noticeable effect on changes in *ElectroDermal Activity (EDA)*. Air pressure demonstrated the greatest contribution towards the detected changes in body temperature and motion. Also, significant correlation was found between air pressure and *heart rate*.

Finally, decision fusion of the classification results from different modalities is performed. To the best of our knowledge this work presents the first attempt at fusing and modelling data from environmental and physiological sources collected from sensors in a real-world setting.

1. Introduction

Repeated exposures to environmental stressors (such as pollution, noise and crowded areas) cause physical illnesses (e.g., headaches, fatigue, sleeping disorder, and heart diseases) and behavioural issues (e.g., stress, attention deficit, anger, and depression) [1–3].

The effect of these stressors on health has been a focal point in health research. Models have been widely used as indispensable tools to assess effects of environmental factors on human and health. In particular, modelling the level of exposures to environmental pollutants such as [4,5].

A decade-long study of 6.6 million people, published in the Lancet recently, found that one in 10 dementia related deaths in people living within 50 m of a busy road was attributable to fumes and noise. There was a linear decline in deaths the further people lived away from heavy traffic [6].

Additionally, Chen’s group [6] noted that because air pollution exposure was estimated at the postal-code level, it may not account accurately for each individual’s exposure. The study suggested that more research to understand this link is needed, particularly into the effects of different aspects of traffic, such as air pollutants and noise at a higher granular levels.

In general, epidemiological and statistical analysis are usually studied based on observed environmental data, which have traditionally been obtained from governmental sources or from a number of sporadically distributed sensing nodes. In both cases, the

* Corresponding author.
E-mail addresses: eiman.kanjo@ntu.ac.uk (E. Kanjo), eman.younas@mu.edu.eg (E.M.G. Younis), Nasser.Sherkat@ntu.ac.uk (N. Sherkat).

http://dx.doi.org/10.1016/j.inffus.2017.05.005

performance of these studies is evaluated against relatively few directly measured data points [7].

Conversely, the capabilities and availability of cheaper, more sensitive and sophisticated sensors for gases, particulates, water quality, noise and other environmental measurements have improved and are enabling researchers to collect data in unprecedented spatial, temporal and contextual detail [7,8].

These sensors range from bespoke devices designed for specific applications, to those found on more mainstream personal devices, such as smartphones. In some cases, people may act as environmental sensors by reporting what they see, hear and feel by participating in the citizen science of environmental conditions [9]. By leveraging widely available wearable devices, communication and sensor technologies many new sensor systems are relatively low-cost compared with technologies used in established monitoring stations [10,11].

Advances in data science and fusion techniques are critical to enable researchers to make best use of the vast amounts of additional, heterogeneous sensor data sources.

Despite the popularity of using wearable sensors for emotion recognition, the problem of quantifying the relation between environmental variables and physiological body reactions and emotions has been overlooked. In addition, the relationship between emotions and all the other environmental and body factors have been studied qualitatively.

In this paper, we incorporate a sensor-data driven approach to understand the relationship of various environmental measures with wellbeing and emotion. By unobtrusively collecting data from on-Body and environmental sensors we can get better understanding of the association and causality of the environmental bases for human health including psychological changes.

This leads us to investigate the following research questions:

1. How can we model and fuse the relationship between on-body and environmental variables?
2. Can the multi heterogeneous sensors integration improve our understanding of the associations and environmental impact on human health?
3. How can information fusion best make use of the 'on-body and environmental Sensor Data' to infer emotion?

Our approach to answer these questions is based on two phase framework in information fusion, which utilizes the new available heterogeneous sensors of multiple modalities as mobile interfaces by studying the relationship between these data sources in spatial-temporal context. Moreover, by studying its relationship with emotion based on decision fusion.

In order to follow our approach, we collected data from forty subjects using on-body sensors 'in the wild' around Nottingham city centre environment. The data collected include on-body data such as body movement, heart rate (*HR*), Electrodermal activities (*EDA*) and body temperature and, environmental data including noise level (Env-noise), air pressure and ambient light levels (*UV*), as shown in Fig. 1.

In addition, collected *GPS* data record the user locations while gathering data. The different data channels are collected, cleaned, aggregated and smoothed for different users and user emotions labels are collected using self-report input, based on 5-step SAM Scale for Valence taken in [12].

The selection of sensors and data analysis techniques is optimized from the ground up with the emotion inference application in mind for outdoor environments.

We have adopted an information fusion approach to analyse and model the data since this method offers an effective solution to many of the issues found in analysing data from individual sensors. Information fusion allows integration of independent features and prior knowledge and, provides a better means of identifying specific aspects of the target application domain and improve robustness against interferences of data sources [13].

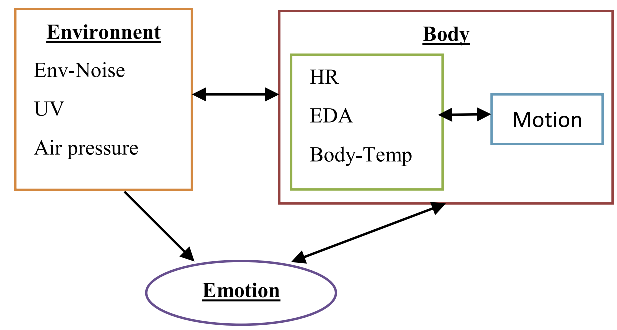

Fig. 1. The relationship between different modalities, the environment, human body, Motion and emotions data.

For examples physiological data, such as heart rate reveal the physical effort of an activity but they may be influenced by external factors such as environmental conditions or social interaction. All of these sources provide only partial information related to the actual individuals' activity.

In this work we utilise, multi-sensor fusion to demonstrate the feasibility of capturing diverse and multi-model derived features in order to identify relationships, associations and causality and, formalize models describing people's reaction and emotions.

Our data fusion approach is in three folds: (1) Data fusion by collecting data from multiple sources including *HR, EDA*, body temperature, movement and activity, environmental noise, location, air pressure and *UV*. (2) Feature fusion by examining relationship between our environmental variables and physiological variables based on exploratory statistics and Multivariate Regression modelling, also by looking at the variable importance and variation (3) Decision fusion by combining multiple classifiers from different modalities for emotion prediction.

The rest of the paper is structured as follows. Section 2 discusses related work focusing on previous efforts in quantifying environmental health impact along with a brief review of on-Body sensors and related information fusion techniques. Section 3 covers the methodology including the user study, system architecture of the proposed method, initial data processing and descriptive statistics. Also Section 4 introduces multivariate regression and its math quotation. Section 5 reports the results of the multimodel analysis and emotion prediction based on decision fusion. Followed by discussion and conclusion sections respectively.

2. Related work

2.1. Quantitative assessment of environmental health impacts

Human exposure to environmental pollutants such as noise, air pollution, traffic or even crowded areas can cause severe health problems ranging from headaches and sleep disturbance and heart diseases [1,2].

The relationship between human body and the environmental factors has been extensively studied in social and environmental sciences, psychology and environmental health literature [3,14]. WHO, defines "Environmental Burden of Disease" [15] as one methodology for quantitatively assessing environmental health impacts at the population level in terms of deaths, Disability Adjusted Life Years (DALYs), or occasionally the number of cases. Other indirect measures can be used to estimate health impacts, for example the number of hospital admissions.

According to WHO quantitative assessments of health impacts are based on combining exposure data with exposure-response information. Such assessments require (i) the compilation of exposure data, (ii) a systematic review of evidence from epidemiology and other scientific disciplines concerning the association between environmental factors and human health, and (iii) the combination of exposure.

For examples, very recently a study has found that in large population-based cohort living close to heavy traffic was associated with a higher incidence of dementia but not with Parkinson's disease or multiple sclerosis [6].

In addition, the negative effect of noise on human health are discussed extensively in the literature including health issues related to sleep disorders, heart problems, vision problems and many more [14].

Similarly, many previously medical studies have confirmed that changes in temperature, humidity, weather events such as storms can trigger asthma attacks [18].

A criticism of statistical epidemiologic models is their focus on identifying association, while causality remains difficult to assess, despite the fact that many information theoretical and physical based models have been developed recently for dissecting spatial-temporal correlation time series more deeply than with traditional statistical models [7].

For examples, the average environmental exposure across regions rarely reveals the specific health problems people face in any given location. Most people live around urban areas. They go down and walk about on city streets and get around by cars, trains or buses. Therefore, in order to know more about the impact of their surrounding and current environment there is a need to monitor people while carrying their daily activities.

For example, cyclists might get exposed to a high level of pollution in half an hour when riding their bikes behind buses than other people get in an entire month. There is a need to monitor and assess people's exposure and health impact in short term and at high granular spatial scale.

Most of the related traditional statistical models do not take advantage of the availability and affordability of modern sensors for on-body and environmental data collection that can make it possible to collect accurate environmental and health data for analytics and modelling.

The increasing pervasiveness of wearable and sensor devices has created new opportunities for sensing people's activities around physical spaces. These new data sources at high level of granularities enables higher level of estimation of human exposures to environmental conditions and quantifying health-related responses that may be associated with such exposures.

Some attempts have been made to use data driven approaches to characterise the impact of environment on health. For examples, mobile phone data have been used to parameterize population movement networks to the spread of malaria [16,17].

Recently, marrying data from personal monitoring devices with air pollution models has improved the characterizations of air pollution exposures [19–21], and in other cases, has employed energy expenditure sensors to improve exposure prediction [22]. Beside health impact, emotions and physiological changes have also started to grab attention as a direct influence on wellbeing. Kööts et al. [23] studied the relationship between positive and negative emotions and the environmental changes such as temperature. Park and Farr [24] studied the relationship between lighting and emotions in a business retail environment. In response to this, we have added the following to the related work section: Gravina and Fortino have developed a novel algorithm designed to detect the (Cardiac Defense Response) CDR by analysing the electrocardiogram (ECG) signal [13]. This approach helps in detecting preceding negative emotional states including fear, chronic worry and panic. This approach helps in detecting preceding negative emotional states including: fear, chronic worry, and Panic.

In addition, many research projects have studied emotion and its relationship with health and physiological changes [25–36,52, 55–66], however none of them have considered integrating physiological and health sensors along with environmental sensors, in order to model and predict emotions.

In this work, we present an emotional analytical model where the environmental and the physiological measures have been combined. Also, we study the relationship between environmental and health variables based on sensor data collected from forty participants walking along the same urban route in Nottingham. Both environmental and physiological data are collected simultaneously along with spatial and temporal information in order to understand at a small scale the relationship between these parameters along with emotion.

2.2. On-body sensors

Body physiological signals require sensors for their measurements. In the past wearable sensors were intrusive and uncomfortable to be used in the real world experiments. However, nowadays with the advancement of the wearable sensors and mobile technologies these sensors have become non-invasive and comfortable for the users, with the availability of wrist-bands, equipped with many built-in sensors. Table 1 presents a list of on-body sensors that have been used for emotion detection:

In addition to the above sensors, currently many wristbands and wearable devices offer a wide range of sensors that are not restricted to health or body statistics. For examples, pollution sensors along with weather stations and other environmental sensors such as light and colours are widely available in different shapes and styles [8,9,19].

Many researchers have started to look at different ways of programming and managing these sensors And, to fuse the data using various computational methods such information fusion [13].

2.3. Information fusion

Information Fusion is the merging of information from heterogeneous sources with differing conceptual, contextual and typographical representations. It can be performed on three levels:

- First, "*Data Level*" fusion aims at collecting different data elements from different sensors to complement each other. It can be done during data collection to fuse external data sources such as user self-reporting of emotions [13,42].
- Second, "*Feature Level*" fusion is performed during data analysis to find the best set of features for the classification. Feature level fusion has been done in [36], to find the best combination of features using EMG, RSP, SC and ECG signals for emotion recognition.
- Third, "*Decision Level*" fusion, which aims to combine the results of multiple techniques to improve decision making. A recent review of various data fusion techniques and applications in body sensor networks can be found in [13]. Granero et al. [42], used feature level fusion to classify emotions and proved that the ECG and *EDA* signals are the most significant signals in emotion classification.

3. Methodology

3.1. Data collection

The data collection setup is depicted in Fig. 2. In this process, we gather various sensor and self-report data from a smart phone application named "*EnvBodySens*", and Microsoft wristband 2. Collected data is then logged and stamped with the time and date. The application also records the data shown in Table 2.

Table 1
List of physiological sensors and signals used widely for emotion detection.

Sensor	Signals and characteristics
Heart Rate	The produced signal is showing the changes in the heartbeats over time. The distance between two consecutive pulse peaks is called the RR interval. It has been widely used in many emotion recognition studies such as [37–42] to measure health and emotions.
Body Temperature	Although, the temperature signal is very simple, it could be used as an indicator of the person's emotions and mood changes [37,33]. Chung et al. [40] proved that the nervous system activity can be detected by changes in skin temperature called Temperature Variability (TV).
Breathing	It has been used widely to measure how fast the person is breathing and patterns of breath. It has proved to be correlated with the heart rate and person's emotions [28,30,36].
Motion	Modern accelerometers include tri-axial micro-electro-mechanical systems (MEMS) for three-dimensional acceleration measurement with sub-second time resolution. However, for analysis, these measurements are usually converted into a uni-axial representation, measuring cumulative activity for a certain period of time. For simplicity, the motion can be represented as the root mean square of all the three components such as $\sqrt{X^2+Y^2+Z^2}$. The accelerometer is now embedded in almost all mobile phones and recently used for emotion recognition in [35].
Electrodermal Activity	Also known as Galvanic Skin Resistance (GSR) has shown high correlation with the emotions and stress detection [29,30,33,34,37,39,42].
EEG Headsets	EEG devices are normally used to measure the electrical activity of the brain. It has been used to measure the emotions and attention [25,41,42].
Muscle contraction (EMG)	EMG measures the electrical pulsed resulted from muscle contraction. It has proved effective in detecting arousal in [30,34,36,38,39,41,42].
Blood Volume pulse (BVP)	PVP has been used for emotion recognition always combined with one or more of the previous sensors [30,39,36].

Fig. 2. (left) Screenshot of EnvBodySens application, (right) Data Collecting process.

Table 2
List of the collected data.

Microsoft band 2	Android phone 6
Heart Rate (*HR*)	Environmental Noise (*Env-Noise*)
Electro Dermal Activities (*EDA*)	GPS Location
Body Temperature (*Body-Temp*)	Self-Report of Emotion (1–5)
Hand Acceleration (Motion)	
Air Pressure	
Light (*UV*)	

In the *EnvBodySens* application, an interface is implemented for continuous and quick labelling of user emotions while walking and collecting data. When the user launches the application, mobile interface appears with five iconic facial expressions ranging from very negative to very positive. A screen capture is presented in Fig. 2 (left) showing the five emotion buttons.

Users are asked to constantly select one of the affective categories (in the form of buttons) as they walk around the city centre. We disabled the screen auto sleep mode on our mobile devices, so the screen was kept on during the data collection process. We adopted the 5-step SAM Scale for Valence taken from [10] to simplify the continuous labelling process.

The dimension ***valence*** fits well into our experimental setup since it describes the positive or negative feeling caused by a situation, an object or an event.

The study was launched in July 2016. A call for participation in the study was advertised in various mailing lists. Forty participants took part in the study all female with an average age of 28. The study was approved by Nottingham Trent University's Ethics Committee. We have chosen to recruit female participants only in order to *rule out* other *factors* (i.e. confounders) related to gender.

Participants were scheduled to take part in the study in order to collect data around Nottingham city centre. The participants have been asked to meet with a trained researcher in a low stress environment (a café) where they were given light refreshments while the experiment is set up. The researcher provided them with details of the study protocol, obtained informed consent, and equipped them with the study equipment. The study was carried out over a number of days in order to find convenient times for the participants and to allow for the limited number of devices available. Information and study details were sent to the participants ahead of the data collection session. The participants were asked to spend no more than 45 min collecting data.

The reasons for limiting the journeys to 45 min are as follows: (1) the shopping route is relatively narrow and can be walked along during this time frame. (2) users from previous experiments found it hard to walk longer.

Based on the previous experience we have found it difficult to motivate participants to walk longer [53,54]. Additionally, we plan

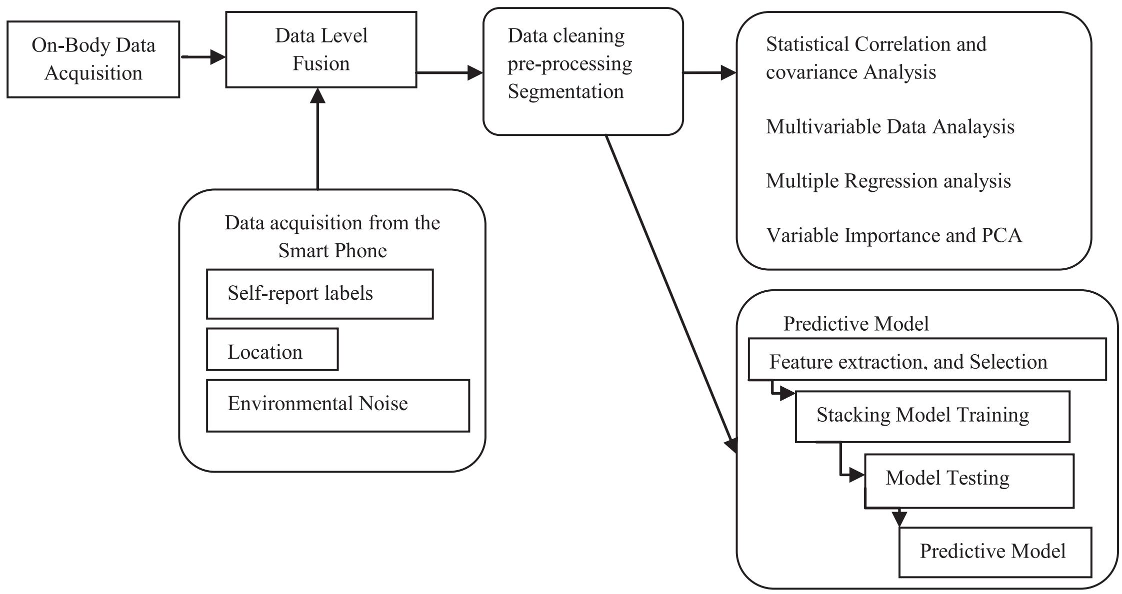

Fig. 3. Overall block scheme of the proposed Information Fusion system (Two phase frame work).

to carry out further analysis on the data that requires adopting fixed route with pre-defined time frame.

Data was collected in similar weather conditions (average 20°), at around 11am. During the data collection process 550,432 data lines were collected as well as 5345 self-report responses.

3.2. System architecture

Our approach consists of real-time collection and off-line analysis of the sensor data using information fusion techniques at all the three fusion levels. The architecture is composed of a number of processing blocks as depicted in Fig. 3. First, the data is collected using on-body sensors and then fused with other contextual user data such as location, noise and emotional state (self-report) which is the data-level fusion. Second, the collected data is cleaned and pre-processed. Third, Statistical correlation, covariance and multiple regression analysis are performed. Fourth, the emotion predictive model is created by extracting features from sensor data and using feature selection for decision fusion. Then, stacking model training is carried out using training examples and then testing the model using unseen data for evaluation.

3.3. Data pre-processing

After the data acquisition the signals are pre-processed and cleaned. Then, the first and the last 30 s were cut from the beginning of the data collection process for each user dataset, the reason for this step is that participants needed a few seconds to fully get into the movement and also few seconds to stop the data collection at the end of the experiment.

Data from six users were excluded due to logging problem, for examples one user was not able to collect data due to battery problem with the mobile phone, another one switched the application off accidently.

We produced Lagged Poincare plots for each individual data subset, in order to remove the ones with abnormal heart rate patterns. The Poincare plot is a visual tool that uses the ratio between standard descriptors for short-term correlation (SD_1) and long-term correlation (SD_2) between *RR* intervals to assess the health of the heart. It has been found that the peculiar shape of *RR* interval is not an artefact or mere placement of point but a specific temporal correlation between the successive *RR* intervals and hence prelates closely to the natural rhythm of heart as a response to many different complex closed loop systems controlling the heart [43]. Given a time series of the form: x_t, x_{t+1}, x_{t+2},..., a return map in its simplest form first plots (x_t,x_{t+1}), then plots (x_{t+1},x_{t+2}), then (x_{t+2}, x_{t+3}), and so on.

The shape of the *RR* interval distribution shows an elliptical pattern and the ratio of SD_1/SD_2 should be higher for a healthy person. Conversely, the shape of *RR* interval distribution is a non-elliptical pattern and the ratio is much lower for a subject with impaired heart or reduced *HRV* [67]. The typical cases of normal and impaired subject are as shown in the right panels of Fig. 4.

HR data from our users have been checked using Poincare plots. All our pariticpants have noramal patterns similar to Fig. 4 (right).

3.4. Statistical analysis

Various statistical methods including descriptive statistics, covariance and correlation matrixes, and Principle Component Analysis (*PCA*) map have been used to identify variables to be included in the multiple regression analyses. Table 3 shows descriptive statistics of the data estimated for all subjects. This includes: the mean (μ), standard deviation (*std*), median, minimum (min), 1st Quartile (*25%*), 2nd Quartile (50%), 3rd Quartile (75%), maximum (100%) and the skewness and kurtosis of the various body and environment sensor signals, where $N = 472{,}904$ samples (after data cleaning).

The correlation matrix in Fig. 5 shows a low level of correlation between the independent variables which suggests that our model will not be affected by the Multi-collinearity problem, which is a basic pre-condition for applying multiple linear regression analysis.

Fig. 6 shows the *PCA* map for all the variables indicating that the first *PCA* component has positive coefficients for all the *on-body* measurements such as *HR, EDA, Motion*, and *Body-Temp*. That is why the three vectors are directed into the right top-quarter of the plot; while all the environmental measurements including *Env-Noise, Air-Pressure, UV* and *Motion* are on the lower half of the plot. Thus, we need to further understand the relationship between both of them and their relation to human emotions.

Based on the covariance matrix, if the covariance is positive, this means that the two variables are mutually increasing. Conversely, if the correlation is negative, this means that the two vari-

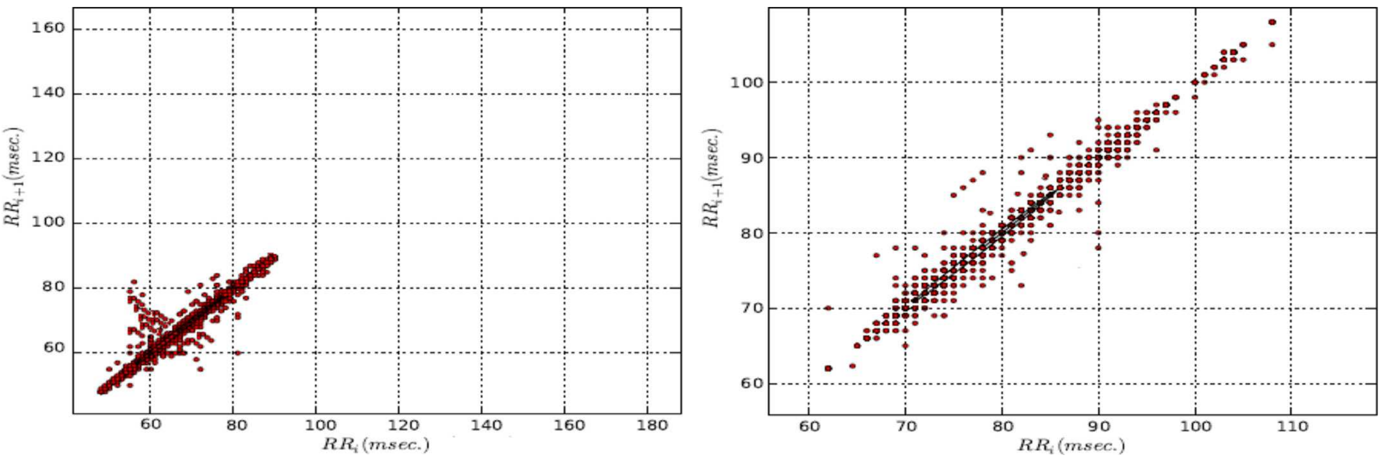

Fig. 4. Abnormal (left) and normal (right) Poincare plots.

Table 3
Descriptive, summary statistics for the collected signals.

	M	σ	Median	Min	1st Qu.	2nd Qu.	3rd Qu.	Max	skewness	kurtosis
Air-Pressure	1014	5.41	1014	1002	1012	1014.0	1019	1020	−0.882	2.81
EDA	1454	348.9	343	15	185	347	952	2903	9.44	90.8
Env-Noise	54.20	3.786	53	20	52	53	55	95	−0.33	14.4
HR	80.24	11.75	77.00	49	73	77	83	189	2.91	18.02
UV	795.9	2646.06	131.0	0.0	47	131	418	62,359	9.9	137.8
X	−0.15	0.662	−0.081	−4.27	−0.82	−0.08	0.28	2.27	0.09	2.16
Y	−0.01	0.625	0.018	−2.77	−0.61	0.018	0.515	3.93	−0.03	1.84
Z	0.01	0.47	0.12744	−1.92	−0.39	0.127	0.348	3.86	−0.27	2.26
Body-Temp	28.93	1.62	28.93	24.67	27.7	28.8	29.89	33.8	0.49	3.46

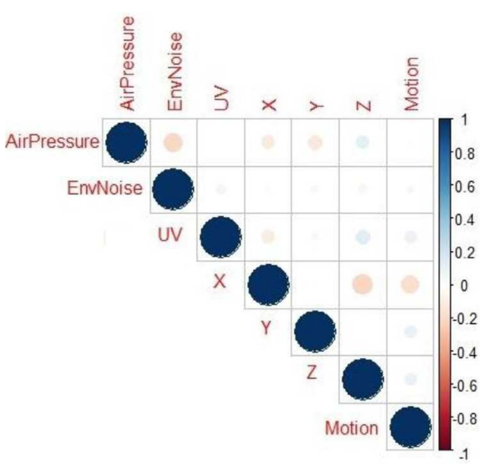

Fig. 5. Correlation Matrix of the independent variables.

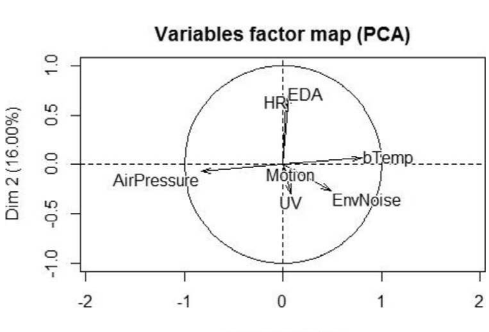

Fig. 6. PCA plot for all the on-body and environmental measurements.

ables are mutually decreasing. If the covariance is zero, this means that there is no relationship between the two variables. It is noticed from Table 4 that the air pressure has negative relation with all the other factors. Whereas. The *EDA* is negatively related with *UV* and *Body-Temp* and positively related with *Env-Noise*, *HR* and *Motion*. In addition, *HR* is negatively related with *UV* and *Body-Temp*. In addition, the Body-Temp is negatively related to *EDA* and positively related with the *Env-Noise, HR, UV* and Motion. In addition, Motion is negatively related with positively related with all the other variables.

Based on the above analysis, we included all the independent variables for the Multivariate Regression analysis in the next section.

It should be noted that EDA and HR have the highest positive correlation with Affect labels as shown in Fig. 7.

4. Results

4.1. Multi-variant regression analysis and variable importance

Having examined our variables closely in order to provide an analytical model we employ multivariate (and multivariable) analysis in order to study the variable dependency between two different modalities using Multivariate Regression and Principle Component Analysis (*PCA*). Statistically speaking, multivariate analysis refers to statistical models that have two or more dependent or outcome variables [44] and multivariable analysis refers to statistical models in which there are multiple independent or response variables.

The analysis is performed in two main steps: First, study the relationship between every dependent variable individually (i.e. the body responses) and all the other independent variables (i.e. the environmental variables). Second, determine the relative importance of the independent variables (i.e. regressors) on each of the

Table 4
The covariance matrix for all the environmental and body measurements.

	Air Pressure	EDA	Env-Noise	HR	UV	Motion	Body-Temp
Air pressure	28.784	−2.674	−4.073	−2.300	−133.653	−0.0165	−4.073
EDA	−267.481	8.239	319.497	3635.712	−1952.3	0.4607	−77.08
Env-Noise	−4.073	3.194	14.386	3.166	574.733	0.0214	0.869
HR	−2.300	3.635	3.166	138.587	414.358	0.0303	0.764
UV	−133.65	−1.952	574.73	−133.65	7070.60	0.0303	38.769
Motion	−0.0165	4.607	0.0214	−0.0165	37.143	37.143	0.0025
Body-Temp	−4.073	−7.708	0.869	−4.073	38.769	0.002	2.51

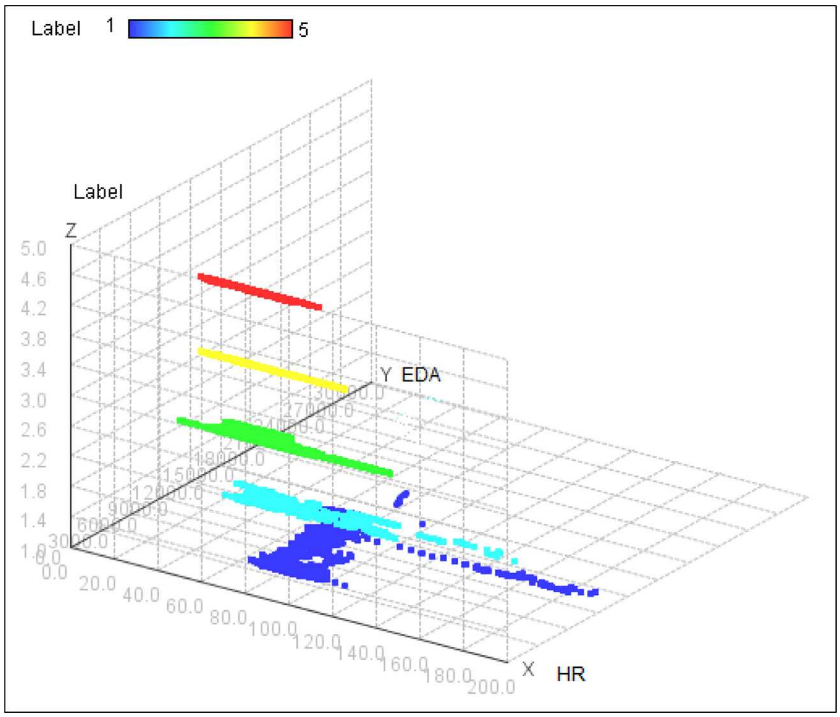

Fig. 7. 3D scatter plot shows how both EDA and HR correlate positively with the label (affect state).

dependent variables using PCA based on the residuals for every dependent variable in the regression model [45].

Multiple Linear Regression analysis was conducted separately for each dependent variable representing body stats in relation to all the independent variables which are represented by the environmental stats. The aim of this step is to determine which body variable can be best predicted using the environmental measurements as independent variables.

Let z_1; z_2;z_3;z_4 be a set of r independent variables (*Env-Noise*, Air pressure and *UV*) believed to be related to a dependent variable Y.

The linear regression model for the 4th sample unit has the form:

$$Y_j = \beta_0 + \beta_1 z_{j1} + \beta_2 z_{j2} + \beta_3 z_{j3} + \epsilon_j, \quad (1)$$

Where, ϵ is a random error and the $\beta_i = 0;\ 1;\ 2;\ 3$ are unknown (and fixed) regression coefficients. $Y_j = 0;\ 1;\ 2;\ 3$, are the four dependent variables (*HR*, *EDA*, *Body-Temp* and Motion respectively) β_0 is the intercept and sometimes we write $\beta_0 z_{j0}$, where $z_{j0} = 1$ for all *j*.

We assume that:

$$E(\epsilon_j) = 0; \quad Var(\epsilon_j) = Q^2; \quad Cov(\epsilon_j, \epsilon_k), = 0 \quad 0\ \forall\ \mathrm{j}\ 6 \neq \mathrm{j}. \quad (2)$$

Then we calculate the residual *e* of the model which is the difference between the observed value of the dependent variable $\hat{y}$ and the estimated value *y*. Each data point has one residual which is expressed as follows:

$$e = \hat{y} - y \quad (3)$$

4.1.1. Multiple regression model for heart rate

The following discussion presents the multiple linear regression model of *HR* using all the other independent variables – environmental factors including (*Env-Noise, Air Pressure and UV*). Table 5 shows multiple regression results for *HR*:

The multiple linear regression model for the heart rate is then formulated using the following equation:

$$Y_j = 211 - 0.0115 z_{j1} - 0.255 z_{j2} + 0.00008 z_{j3} + \epsilon_j \quad (4)$$

The comprehensive model above was evaluated using the diagnostic regression curves shown in Fig. 8 shows the relation between the fitted values against the model residual values (i.e. goodness of fit). The model is statistically significant based on (p <0.001).

The right *Q-Q* plot in Fig. 8 shows that data exhibits a pronounced bimodal distribution, which may be seen clearly in the left residual plot. Normal *Q-Q* plots constructed from bimodal data typically exhibit a ‘twist’ like the one seen in this plot. To explain this behaviour (of why the upper part of the plot looks deviated from the baseline) the lower portion of the *Q-Q* plot is almost linear, suggesting an approximate normal distribution, corresponding to one mode of data distribution. Similarly, the upper portion of

Table 5
Multiple regression analysis between HR (dependent variable) and relevant independent variables.

Independent variable	Regression coefficient (β)	Std. error	*t*-value	*P*-value
Intercept	211	3.86	54.7	<2e−16***
Air-Pressure	−0.0115	0.0037	−30.9	<2e−16***
Env-Noise	−0.255	0.0053	−48.26	<2e−16***
UV	0.000077	0.0007	10.4	<2e−16***

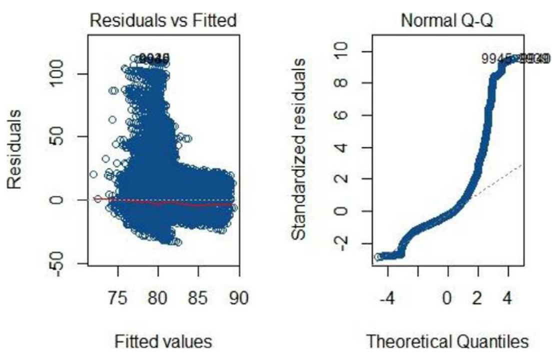

Fig. 8. HR Diagnostic regression curves: (Left) represents residuals curve, and (Right) represents the Q-Q curve.

Table 6
Variable importance for HR.

Aggressor/Metrics	Img	Last	First	Pratt
Air Pressure	**0.2496**	**0.2818**	**0.0867178**	**0.2031**
Env-Noise	**0.72066**	**0.6861**	**0.7700**	**0.73090**
UV	0.0489202	0.0526879	0.0444603	0.0479666

the *Q-Q* plot is again roughly linear, but with a very different intercept that corresponds to the larger mean of the data distribution (i.e. the duration of the small peaks in environmental changes). To connect these two 'roughly linear' local segments, the curve must exhibit a rapid transition region between them (i.e. the duration of the large peaks in environmental changes). By the same reasoning more general multi-modal distributions will exhibit more than one such "twist" in their *Q-Q* plots.

4.1.2. Variable importance for HR

The variable importance calculations measures produce a predictor ranking (also known as variable importance) based on the contribution predictors make to the construction and variability of the model [46]. Table 6 shows the four *HR* variable importance metrics for each of the independent variables.

The metrics in bold font indicate that *Air-Pressure* and *Env-Noise* affect *HR*, whereas, the *UV* metrics has less effect on the *HR*. Furthermore, adding the motion as an independent variable to the *HR* model does not make any visible changes to the model. The ANOVA test [45] applied on the two models indicates that there is no difference between the two models in terms of the residuals.

To conclude, the heart rate (*HR*) is affected by the *Env-Noise* level (the most important variable in our model), with the *Air-Pressure* in the second place. However, the *UV* and Motion, have proven to have no significant effect on the heart rate.

These initial findings are in agreement with Scientists who have now shown that exposure to noise during everyday life influences heart rate variability. Many previous works which suggest a direct impact of high level of irregular noise levels on the regular rhythm of the heart. For example recent studies have found that noise levels between <55 and >75 dB are linked to heart related diseases such as coronary heart disease [55]. Another study shows that HRV

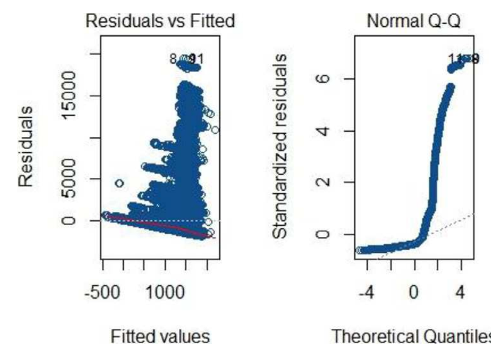

Fig. 9. EDA Diagnostic regression curves: (Left) represents residuals curve, and (Right) represents the Q-Q curve.

was affected in association with increases of 5 dB in noise exposure at both the higher and lower noise level ranges. The study showed that not only higher noise levels have a stressful effect and are harmful to health, but that lower noise levels can cause adverse health effects too [56].

However, the impact of air pressure is still subject to debate. For example a study published by the American Heart Association showed that atmospheric pressures increased an individual's risk of heart attack [58]. While another study has examined the links between atmospheric conditions, temperature and air pressure with the occurrence of various cardiovascular events, they have not found enough evidence to suggest a direct impact of air pressure on cardiovascular events [57].

4.1.3. Electrodermal activity (EDA)

Table 7 shows multiple regression results for *EDA*:

The multiple linear regression model for the *EDA* is then formulated using the following equation:

$$Y_j = 6771.45 - 6.37 z_{j1} + 21.58 z_{j2} - 0.029\ z_{j3} + \epsilon_j, \quad (5)$$

The comprehensive model above is evaluated using the diagnostic curves. The following is the *Q-Q* plot and the residuals of the final linear equation. The model is statistically significant ($p < 0.001$).

Similar to the *HR Q-Q* plot, *EDA Q-Q* plot on the right of Fig. 9 shows, that data exhibit a pronounced bimodal distribution which may be seen clearly in the left residual plot. The lower portion of the *Q-Q* plot is almost linear suggesting an approximate normal distribution corresponding to one mode of data distribution. Similarly, the upper portion of the *Q-Q* plot is again roughly linear but with a much different intercept that corresponds to the second mode in the data distribution.

4.1.4. Variable importance for EDA

Variable Importance metrics show that, the *Env-Noise* level and the *UV* both have a similar effect on the *EDA* while the *Air-Pressure* has a less significant effect on the *EDA*. Moreover, adding the motion as an independent variable to the previous model is statistically significant ($p < 2.2e-1$) (Table 8).

Table 7
Multiple regression analysis between EDA (dependent variable) and relevant independent variables.

Independent variable	Regression coefficient (β)	Std. error	t -value	P-value
Intercept	6771.45	944.77	7.16	7.66e−13***
Air-Pressure	−6.37	0.91	−6.96	3.34e−12***
Env-Noise	21.58	1.29	16.63	<2e−16***
UV	−0.029	−0.001	−16.26	<2e−16***

Table 8
Variable importance for EDA.

Aggressor/Metrics	Img	Last	First	Pratt
Air Pressure	0.1265742	0.082174	0.1659	0.11874
Env-Noise	**0.4712**	**0.4691**	**0.47382**	**0.48014**
UV	**0.40220**	**0.448646**	**0.36019**	**0.4011**

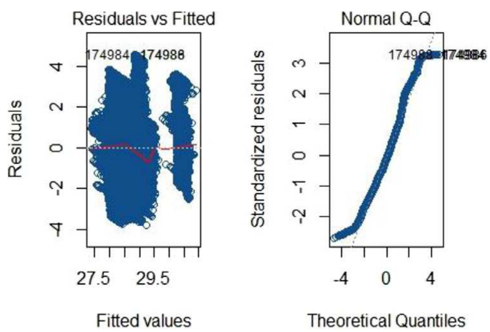

Fig. 10. Body-Temp Diagnostic regression curve: (Left) represents residual curve, and (Right) represents the Q-Q curve.

4.1.5. Body temperature

Table 9 shows the results for the multiple regression for *Body-Temp* as dependent variable.

The multiple linear regression model for the skin temperature is then formulated using the following equation:

$$Y_j = 168 - 0.014 z_{j1} - 0.0211 z_{j2} - 0.0000001\ z_{j3} + \epsilon_j, \qquad (6)$$

The model is statistically significant ($p < 0.001$). Fig. 10 shows the *Q-Q* plot for the residual and the residual against the fitted values of the final linear equation.

The *Q-Q* plot looks perfectly linear and matching the baseline. This indicates that residuals are distributed approximately in a normal fashion. In particular, the residual tend to be larger in magnitude than what we would expect from the normal distribution. Body temperature scored much higher in terms of *R*2 goodness of fit measure 0.35 whereas *HR and EDA* were most difficult to predict using the environmental factors only (Tables 10 and 11).

4.1.6. Variable importance for Body-Temp

The variable importance for *Body-Temp* suggests that the air pressure and the noise are the most effective measures in the model. These values also reveal that the *UV* variable is not effective in this model so it can be removed from the model.

4.1.7. Body Motion

The fourth dependent variable that will be examined here is the motion. The following discussion presents the multiple linear regression model of motion with all the other independent variables (i.e. *Env-Noise, Air pressure* and *UV*).

The multiple linear regression model for the motion is then formulated using the following equation:

$$Y_j = 1.35 - 0.000038 z_{j1} + 0.0012 z_{j2} + 0.0000005 z_{j3} + \epsilon_j, \qquad (7)$$

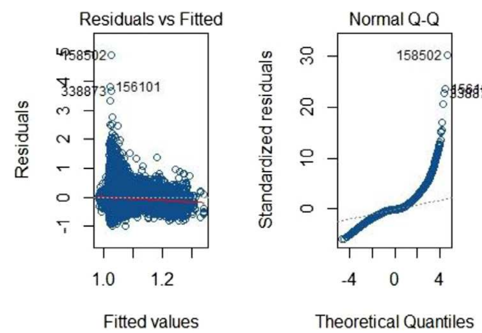

Fig. 11. Motion Diagnostic regression curves: (Left) represents residual curve, and (Right) represents the Q-Q curve.

The comprehensive model above is evaluated using the diagnostic curves. Fig. 11 shows the *Q-Q* plot for the residual and the residuals versus fitted values of the final linear equation.

The *Q-Q* plot looks deviated from the baseline, on the right side, but on the left sides of the baseline, the actual data points are clearly linear, which suggests multi-modality in the data. In other words, the upper part *Q-Q* plot is again roughly linear but with a much different intercept that corresponds to the larger mean of the second peak in the distribution.

4.1.8. Variable importance for motion

The variable importance for body Motion in Table 12 suggests that, the *Air Pressure* and the *Env-Noise* are the most effective measures in the Motion model. These values also reveal that the *UV* variable is not effective in this model.

4.1.9. PCA analysis

The second step in the multivariable data analysis is computing the *PCA* between all the independent variables residuals, to see if there is any additional inter relationships. Our dependent variables are the *HR, EDA*, Motion and the *Body-Temp*. The purpose of the *PCA* is to discover the inter-relationships between the residuals of the models created for these variables previously. Table 13 shows the principle components for the independent variable residuals.

This first component represents almost 80% of data variability indicates that, the *HR* and *EDA* could be used alone to represent the dataset variability. Whereas, the second component suggest that the body temperature alone can represent more than 90% of the variability in the data. The first component is strongly correlated with both the *HR* and *EDA*, whereas, the second component is strongly correlated with the body temperature. The third component suggests that the *HR* and motion are both sufficient to describe the variability in the dataset. The fourth component suggests that the *HR* and *EDA* are the most important variables. Fig. 12 shows the relationship between the first and the second principle components.

4.2. Emotion predictive model

In this section, we will present our approach to model emotion based on our collected data along with performance evaluation.

Table 9
Multiple regression analysis between Body-Temp (dependent variable) and relevant independent variables.

Independent variable	Regression coefficient (β)	Std. error	t-value	P- value
Intercept	168	0.04577	367.3	<2e-16***
Air–Pressure	−0.014	0.0004	−312.2	<2e-16***
Env-Noise	0.0211	0.0006	33.6	<2e-16***
UV	0.000001	0.00008	1.30	0.19207

Table 10
Variable importance for Body-Temp.

Aggressor/Method	lmg	Last	First	Pratt
Air Pressure	**0.94**	**0.98**	**0.91**	**0.96**
Env-Noise	0.05	0.011	0.083	0.031
UV	0.0001	0.0000017	0.00033	0.00007

Table 12
Variable importance for motion.

Aggressor/Method	lmg	Last	First	Pratt
Air Pressure	**0.95**	**0.98**	**0.92**	**0.97**
Env-Noise	0.05	0.011	0.08	0.031
UV	0.0001	0.000018	0.00033	0.00007

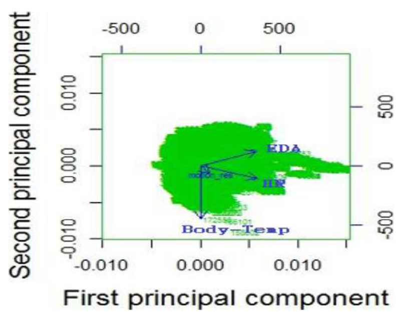

Fig. 12. Biplot of the first two principal components of the PCA.

Table 13
Shows dependent variables' residuals for PCA.

Variable	PC1	PC2	PC3	PC4
*HR*_residuals	0.70725	−0.218	0.0603	0.6698
*EDA*_residuals	0.69719	0.256	0.1012	−0.6621
Body-Temp residuals	0.00918	−0.932	0.1539	−0.3268
Motion residuals	0.11682	−0.133	−0.9810	−0.0784

4.2.1. Feature extraction

After thoroughly analysing the related literature about feature extraction from physiological signals, we also extracted statistical features from the environmental sensors. In total, we extracted 87 features. Our extracted features are as follows:

1. For *HR, EDA* and *Body-Temp* signals, common statistical features were computed: mean, median, max, min, max-min, and standard deviation and, quartiles [33,40].
2. Additionally, for the *HR*:
   Standard *HRV* analysis refers to the extraction of parameters defined in the time and frequency domains [34,36]. In total we extracted 17 *HRV* features.
   Concerning the time domain analysis, we calculated the following: the maximum and minimum of Heart Rate, the square root of the mean of the sum of the squares of differences between subsequent NN intervals $RMSSD = \sqrt{\frac{1}{N-1}\sum_{j=1}^{N-1}(RR_{j+1}-RR_J)^2}$, $pNN50 = \frac{NN50}{N-1}100\%$ *and* $pNN30 = \frac{NN30}{N-1}100\%$ (percentage of consecutive NN intervals which differ by more than 50 and 30 ms respectively), Standard Deviation of the *NN* interval *SDNN. RMSSD* indicates the short-term variability; instead *SDNN* and *HRV* triangular index are features of the entire *HRV* [43].
   We also derive frequency domain features which are indicative of sympathetic and parasympathetic neural activity including: the total spectral power of the successive difference of *NN* intervals in power bands up to *0.04* Hz, between *0.04* and 0.*15* Hz, and between *0.15* and *0.50* Hz, and ratio of low *lf* to high frequency power *hf*.
4. For the *EDA*:
   Additionally, ten features were extracted from the *EDA* signal, including, the number of responses, the power of responses, the number of significant responses (responses which have a value over some threshold) and the power of significant responses and slope and intercept of signal were calculated.
5. For *Motion*:
   We abstracted the motion representation to be one component and it is represented as in [37]:

$$\text{Motion} = \sqrt{X^2 + Y^2 + Z^2} \tag{8}$$

4.2.2. Features fusion level

Our system utilizes feature-level fusion, where feature sets from different modalities are concatenated to form two feature spaces, the environmental and on-Body modalities. As explained in the previous section (section number) *84* features were extracted. However, many of these features do not have an important explanatory effect on the emotional outcomes. In addition, many of the extracted features of the same signal are correlated with each other's and hence can be removed to simplify the model by reducing it to only include the most significant features necessary to explain the emotion response.

Table 11
Multiple regression analysis between Motion (dependent variable) and relevant independent variables.

Independent variables	Regression coefficient (β)	Std. error	t-value	P value
Intercept	1.35	0.0534	25.2	<2e−16***
Air Pressure	−0.000038	0.000005	−7.4	<2e−16***
Env-Noise	0.0012	0.0000007	16.07	<2e−16***
UV	0.0000005	0.0000001	50.2	<2e−16***

We developed a predictive model, to test whether it is possible to accurately predict individual's affect state based on both a combination of physiological and environmental features.

Our labelled data has *5* classes ranging from *Class* $_{\text{very negative}} = 1$ to *Class* $_{\textit{very positive}} = 5$ with *355,089* instances.

To build the model, we tested the levels of significance of the features in relation to the affect lables and checked the response of the affect labels for any interdependency between the variables based on the correlation matrix.

We checked the pairwise correlations between features and the label on the whole dataset. Based on the result of features evaluation, we finally selected 21 features, which have strong correlations with label to build the prediction model (i.e., feature selection step).

4.2.3. Predictive model for emotion recognition based on multimodal fusion

We opt to use a multi-learner approach based on Ensemble algorithm called Stacking [47]. There are several reasons for preferring a multi-classifier system to a single classifier. It is mainly done to improve the accuracy and efficiency of the classification system and the volume of the data to be analysed is too large to be handled by a single classifier. Training a classifier with such a large amount of data is usually not practical. Finding a single classifier to work well for all test data is difficult. Instead multiple classifiers can be combined to give a better output than a single classifier. It may not necessarily out-perform a single best classifier but the accuracy will be on average better than all the individual classifiers.

Stacking model: Stacked generalization (or stacking) is an ensemble learner that combines multiple models. Unlike bagging [48] and boosting [49] stacking is used to combine models of different types.

Stacking exploits this prior belief further by using performance on the testing data to combine the models rather than choose among them, thereby typically getting a better performance than any single one of the trained models. It has been successfully used on both supervised learning tasks (e.g. regression) and unsupervised learning (e.g. density estimation) [47].

Due to the multi-model nature of our features, we follow the stacking approach in [51], in which each modality is processed independently by the corresponding classifier and the outputs of the classifiers are combined to yield the final result instead of concatenating the features to form a composite feature vector and then input to a classifier.

Our procedure is as follows: Let D_1 and D_2 be two different datasets: Environmental D_1 (*including Env-Noise, Air Pressure* and *UV*) and Physiological D_2 including (*HR, EDA, Body-Temp* and *Motion*).

The datasets are then split up into three parts each (D_i^0 to D_i^2), the level-0 training sets, level-0 evaluation sets and level-1 evaluation sets. The classifiers $e_i \in E$ with $|E| = N$ are trained on D_i^0 and evaluated on D_i^1 to produce D_i' the *level-1* training-set parts which are combined to form D' the full *level-1* training-set on which a *level-1* classifier is then trained. The whole stack is then evaluated on the D_i^2 datasets.

In order to train and model our labelled dataset, we stacked a combination of three base classifiers: Support Vector Machine (*SVM*), Random Forest (*RF*) and K Nearest Neighbour (*KNN*); and Naive Bayes (*NB*) as the Stacking Model Learner. We have chosen these classifiers since the have proven to be effective in classifying emotions based on on-body sensors [25–36], and all can output a confidence rating for each label attribute. Our class attribute is of nominal value ranging from 1 to 5. Fig. 13 represents the *Accuracy levels* and *F-measures* of all the base learner models on two modalities and the overall Stacking model. It is clear that Stacking model with five classifiers yields excellent results and outperforms the individual classifiers with *F-measure 0.84* and *Accuracy %86*. It is difficult to compare precisely our results to previous work in the literature, since no other work included environmental sensor data in the emotion model.

The results show the improvement in the classification accuracy of emotion prediction method by combining decision fusion and feature fusion based on Stacking Learner. Furthermore, Illustration of the confusion matrix of the 5 labels is indicated in Fig. 14.

Although our system was developed based on on-body sensor data as well as environmental signals obtained from multiple sensors, the ratio of correct recognition was comparable with that of the previous systems [27–33].

In order to learn more about the influence of each single modality to the overall performance of the prediction, we show the prediction accuracy of each modality and for each user, see Fig. 15.

The line charts indicate high variability in all the two modalities. The most extreme difference occurs in the environmental data. The physiological modality displays better accuracy levels among most of the participants. There are also no suggestions of a correlation between the two modalities, e.g. high accuracy levels of environmental data don't indicate a corresponding high levels in the physiological data. Also the accuracy levels among users vary in random fashion.

5. Discussion

We developed five generalized multiple Regression models to analyse the relative impact of environmental factors on body dynamics. The obtained results quantitatively indicated a possible control of ambient environment factors on body and emotion variables. Individual variables are not significant on their own but they have a significant impact when combined with other independent variables.

Multiple regression results suggest that the *HR* data exhibits a pronounced bimodal distribution. Also it shows that *Air-Pressure* and *Env-Noise* contribute to a large percentage of variation in *HR*, with *Env-Noise* being the most important variable that explains the majority of changes, whereas, the *UV* has much less significance on the *HR* data. In addition, adding the motion as an independent variable to the *HR* model does not make any noticeable changes to the model.

These results comply. Also, these findings are in agreement with previous epidemiologic research concluding that noise exposure can contribute to the prevalence of cardiovascular disease [50].

Similarly, Multiple regression model of *EDA* exhibits bimodal and variable importance metrics showing that the *Env-Noise* level and the *UV* both have a similar effect on the *EDA* whereas the *Air-Pressure* and Motion have a less significant effect.

In addition, the multiple regression model of *Body-Temp* between the independent variables showed a perfect linear model matching the regression base line. The variable importance for *Body-Temp* suggests that, the *Air-Pressure* and the *Env-Noise* are the most effective measures in the model. These values also reveal that the *UV* variable does not have a significant effective in this model.

Furthermore, analysis of the multiple regression model of *Motion* suggests multi-modality distribution in the data, with *Air-Pressure* is the main noticeable relevant variable.

Body-Temp scored much higher in terms of R^2 goodness of fit measure = 0.35 whereas *HR and EDA* were more challenging to predict using the environmental factors alone.

PCA analysis suggests that all the variables used can describe the variability of the data. The *PCA* ordination map suggested that *EDA, HR* and *Body-Temp* are grouped and oriented in one direction, while the environmental variables are oriented towards another direction of the map. *Motion* sits in between moving in different direction from both modalities.

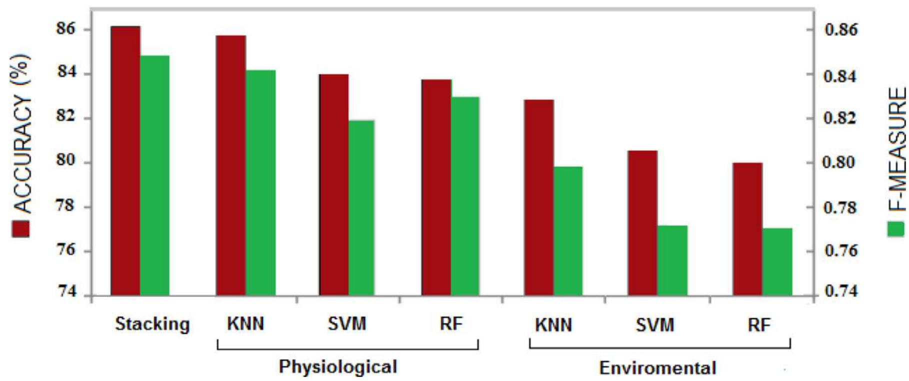

Fig. 13. Accuracy and F-Measure levels of the base learners and the Stacking learner.

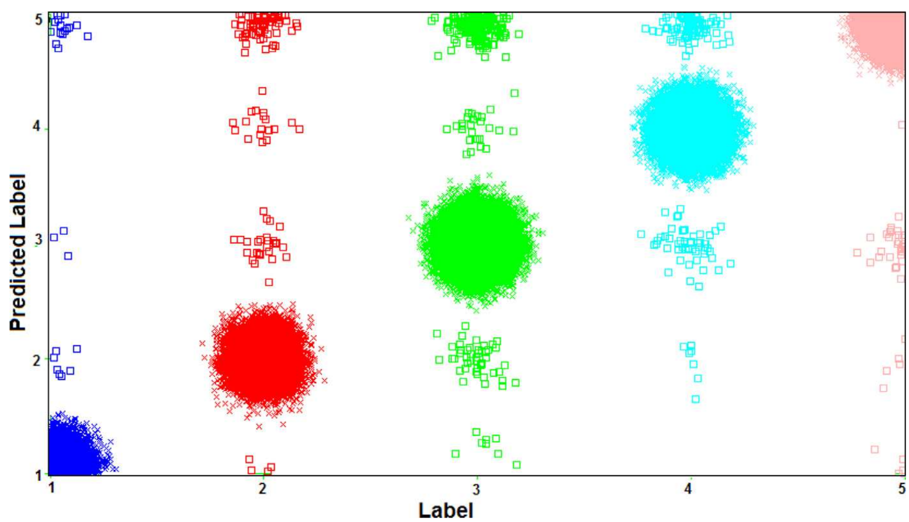

Fig. 14. Visualisation of the Confusion Matrix for the Stacking Learner.

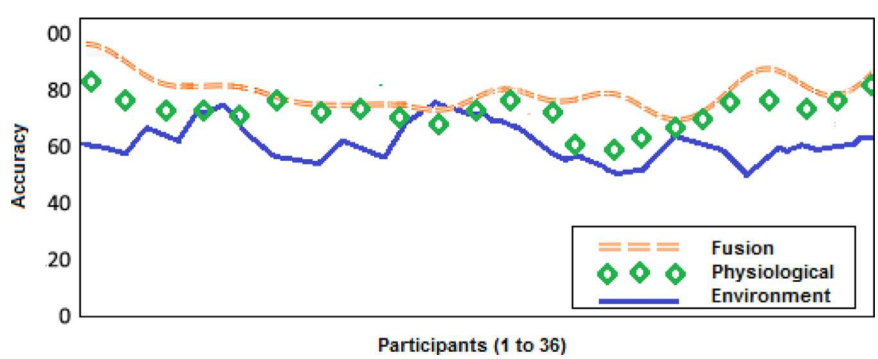

Fig. 15. Prediction Accuracy per user and modality and the fusion approach.

Although we can conclude that various environmental factors can contribute to the prevalence of physiological changes including heart rate variability and body temperature, the evidence for this relationship is still inconclusive because of the limitations in the number of the effectors measured and the exposure characterization.

On the other hand, despite of the quality of the emotion prediction results, it is still difficult to single out the impact of the individual environmental factors compared to the individual physiological measurements. Also, we can't rule out the hidden impact of other environmental co-founders such as gas pollution or crowd size around the street.

Since this is the first set of experiments of its kind, it is hard to compare our results to any other studies based on sensor data feed. As mentioned previously quantifying and modelling the relationships between all these variables haven't been studied before.

Future preventive health strategies should involve environmental and urban interventions. Decision makers have the responsibility to develop, implement, evaluate, and improve guidelines and standards to protect public health around urban spaces; new tools and strategies based on local conditions will have to be developed.

6. Conclusions and future work

In this paper, we have described our information fusion approach for on-body and environmental sensing that offers new opportunities for data-intensive modelling particularly involving the quantification of some aspects of physiological and movement changes in relation to the variation in environmental factors measured continuously in the same Spatial-Temporal context.

To achieve this, we have conducted a real-world study 'in the wild' with on-body and mobile sensors. Data was collected from forty participants walking around Nottingham city centre.

Multivariate linear regression models for on-body sensor data were developed. We found that the spatial variability in on-body sensor data were directly associated with environmental changes. Emotion prediction has resulted in an encouraging accuracy level which is comparable with that of the previous systems. In addition, decision fusion of emotion recognition based on the two modalities yielded an increase in the performance over each single modality, indicating at least some complementarity to the modalities.

These results show that, the realisation of user independent emotion recognition based on the integration of physiological and environmental signals is feasible.

Since we can only collect a limited number of signals the constraints imposed by the on-body instrumentation heavily influence the design of the algorithm. Future work will look at adding more sensor modalities to increase further our understanding of the hidden connections between the environment and health. Also, in future work we will look at modelling these parameters in relation to changes in physical places by aggregating the data into different spatial segments.

Acknowledgements

This work was supported by the B11 Research unit at NTU, Ref 3452. Also, the authors would like to thank Dr. Caroline Langensiepen for her help with the Ethical Approval Application and Dr. Ahmed Aldabbagh for his help in collecting the data.

References

[1] D.W. Dockery, C.A. Pope, Acute respiratory effects of particulate air pollution, Annu. Rev. Public. Health. 15 (1) (1994) 107–132.
[2] D. Briggs, Environmental pollution and the global burden of disease, Br. Med. Bull. 68 (1) (2003) 1–24.
[3] Y.H. Lim, H. Kim, J.H. Kim, S. Bae, H.Y. Park, Y.C. Hong, Air pollution and symptoms of depression in elderly adults, Environ. Health. Perspect. 120 (7) (2012) 1023.
[4] S.S. Lim, T. Vos, A.D. Flaxman, G. Danaei, K. Shibuya, H. Adair-Rohani ...., M. Aryee, A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010, Lancet 380 (9859) (2013) 2224–2260.
[5] U. Schlink, K. Strebel, M. Loos, R. Tuchscherer, M. Richter, T. Lange, A. Ragas, Evaluation of human mobility models, for exposure to air pollutants, Sci. Total Environ. 408 (18) (2010) 3918–3930.
[6] H. Chen, J.C. Kwong, R. Copes, et al., Living near major roads and the incidence of dementia, Parkinson's disease, and multiple sclerosis: a population-based cohort study, Lancet (2016). published online Jan 4. http://dx.doi.org/10.1016/S0140-6736(16)32399-6 .
[7] S. Reis, E. Seto, A. Northcross, N.W. Quinn, M. Convertino, R.L. Jones, M.C. Wimberly, Integrating modelling and smart sensors for environmental and human health, Environ. Model. Softw. 74 (2015) 238–246.
[8] E. Kanjo, S. Benford, M. Paxton, A. Chamberlain, D.S. Fraser, D. Woodgate, A. Woolard, MobGeSen: facilitating personal geosensor data collection and visualization using mobile phones, Pers. Ubiquitous Comput. 12 (8) (2008) 599–607.
[9] E. Kanjo, J. Bacon, D. Roberts, P. Landshoff, MobSens: making smart phones smarter, IEEE Pervasive Comput. 8 (4) (2009) 50–57.
[10] E. Banzhaf, F. de la Barrera, A. Kindler, S. Reyes-Paecke, U. Schlink, J. Welz, S. Kabisch, A conceptual framework for integrated analysis of environmental quality and quality of life, Ecol. Indic. 45 (2014) 664–668.
[11] S. Galelli, G.B. Humphrey, H.R. Maier, A. Castelletti, G.C. Dandy, M.S. Gibbs, An evaluation framework for input variable selection algorithms for environmental data-driven models, Environ. Modell. Softw. 62 (2014) 33–51.
[12] M.M. Bradley, P.J. Lang, Measuring emotion: the self-assessment manikin and the semantic differential, J. Behav. Ther. Exp. Psychiatry 25 (1) (1994) 49–59.
[13] R. Gravina, P. Alinia, H. Ghasemzadeh, G. Fortino, Multi-sensor fusion in body sensor networks: state-of-the-art and research challenges, Inf. Fusion 35 (2017) 68–80.
[14] S. Stansfeld, M. Haines, B. Brown, Noise and health in the urban environment, Rev. Environ. Health 15 (1-2) (2000) 43–82.
[15] World Health Organisation (WHO), 2017 [accessed 09/01/2017]. http://www.who.int/heli/tools/quantassess/en/.
[16] A.L. Barabasi, The origin of bursts and heavy tails in human dynamics, Nature 435 (7039) (2005) 207–211.
[17] A. Wesolowski, N. Eagle, A.J. Tatem, D.L. Smith, A.M. Noor, R.W. Snow, C.O. Buckee, Quantifying the impact of human mobility on malaria, Science 338 (6104) (2012) 267–270.
[18] M. Kampa, E. Castanas, Human health effects of air pollution, in: 4th International Workshop on Biomonitoring of Atmospheric Pollution (With Emphasis on Trace Elements), 2008, pp. 362–367.
[19] S. Steinle, S. Reis, C. Sabel, Quantifying human exposure to air pollution – moving from static monitoring to spatio-temporally resolved personal exposure assessment, Sci. Total Environ. 443 (2013) 184–193.
[20] J. Engel-Cox, T.K.O. Nguyen, A. van Donkelaar, R.V. Martin, E. Zell, Toward the next generation of air quality monitoring: particulate matter, Atmos. Environ. 80 (2013) 584–590.
[21] S. Steinle, S. Reis, C. Sabel, S. Semple, M.M. Twigg, C.F. Braban, A.E. Leeson, M.R. Heal, D. Harrison, C. Lin, H. Wu, Application of a low-cost method to quantify human exposure to ambient particulate matter concentrations across a wide range of microenvironments, Sci. Total Environ. 508 (2015) 383–394.
[22] A. de Nazelle, E. Seto, D. Donaire-Gonzalez, M. Mendez, J. Matamala, M.J. Nieuwenhuijsen, M. Jerrett, Improving estimates of air pollution exposure through ubiquitous sensing technologies, Environ. Pollut. 176 (2013) 92–99.
[23] L. Kööts, A. Realo, J. Allik, The influence of the weather on affective experience, J. Individual Differences (2011).
[24] N.K. Park, C.A. Farr, The effects of lighting on consumers' emotions and behavioral intentions in a retail environment: a cross-cultural comparison, J. Interior Des. 33 (1) (2007) 17–32.
[25] R.B. Knapp, J. Kim, E. André, Physiological signals and their use in augmenting emotion recognition for human–machine interaction, in: Emotion-Oriented Systems, Springer, Berlin Heidelberg, 2011, pp. 133–159.
[26] S.D. Kreibig, Autonomic nervous system activity in emotion: a review, Biol. Psychol. 84 (3) (2010) 394–421.
[27] R.W. Picard, E. Vyzas, J. Healey, Toward machine emotional intelligence: analysis of affective physiological state, IEEE Trans. Pattern Anal. Mach. Intell. 23 (10) (2001) 1175–1191.
[28] E.M. Younis, Sentiment analysis and text mining for social media microblogs using open source tools: an empirical study, Int. J. Comput. Appl. 112 (5) (2015).
[29] M. Valstar, J. Gratch, B. Schuller, F. Ringeval, D. Lalanne, M.T. Torres ..., & M. Pantic (2016). AVEC 2016-Depression, Mood, and Emotion Recognition Workshop and Challenge. arXiv preprint arXiv:1605.01600.
[30] J. Kim, E. André, Emotion recognition based on physiological changes in music listening, Pattern Anal. Mach. Intell., IEEE Trans. 30 (12) (2008) 2067–2083.
[31] B. Guthier, R. Dörner, H.P. Martinez, Affective computing in games, in: Entertainment Computing and Serious Games, Springer International Publishing, 2016, pp. 402–441.
[32] D. Datcu, L. Rothkrantz, Multimodal recognition of emotions in car environments, DCI&I 2009 (2009).
[33] C.L. Lisetti, F. Nasoz, Using noninvasive wearable computers to recognize human emotions from physiological signals, EURASIP J. Adv. Signal Process. 2004 (11) (2004) 1–16.
[34] K. Takahashi, Remarks on SVM-based emotion recognition from multi-modal bio-potential signals, in: Robot and Human Interactive Communication, 2004. ROMAN 2004. 13th IEEE International Workshop on, IEEE, 2004, September, pp. 95–100.
[35] M. Irrgang, H. Egermann, From motion to emotion: accelerometer data predict subjective experience of music, PLoS One 11 (7) (2016) e0154360.
[36] Z. Guendil, Z. Lachiri, C. Maaoui, A. Pruski, Multiresolution framework for emotion sensing in physiological signals, in: Advanced Technologies for Signal and Image Processing (ATSIP), 2016 2nd International Conference on, IEEE, 2016, March, pp. 793–797.
[37] M. Adibuzzaman, N. Jain, N. Steinhafel, M. Haque, F. Ahmed, S. Ahamed, R. Love, In situ affect detection in mobile devices: a multimodal approach for advertisement using social network, SIGAPP Appl. Comput. Rev. 13 (December (4)) (2013) 67–77, doi:10.1145/2577554.2577562.
[38] W. Wan-Hui, Q. Yu-Hui, L. Guang-Yuan, Electrocardiography recording, feature extraction and classification for emotion recognition, Computer Science and Information Engineering, 2009 WRI World Congress on, 4, IEEE, 2009.
[39] K.H. Kim, S.W. Bang, S.R. Kim, Emotion recognition system using short-term monitoring of physiological signals, Med. Biol. Eng. Comput. 42 (3) (2004) 419–427.
[40] W.Y. Chung, S. Bhardwaj, A. Punvar, D.S. Lee, R. Myllylae, A fusion health monitoring using ECG and accelerometer sensors for elderly persons at home, in: 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2007, August, pp. 3818–3821.
[41] N. Ramzan, S. Palke, T. Cuntz, R. Gibson, A. Amira, Emotion recognition by physiological signals, Electr. Imaging 2016 (16) (2016) 1–6.
[42] A.C. Granero, F. Fuentes-Hurtado, V.N. Ornedo, J.G. Provinciale, J.M. Ausín, M.A. Raya, A comparison of physiological signal analysis techniques and classifiers for automatic emotional evaluation of audiovisual contents, Front. Comput. Neurosci. 10 (2016).

[43] A.H. Khandoker, C. Karmakar, M. Brennan, M. Palaniswami, A. Voss, Poincaré Plot Methods For Heart Rate Variability Analysis, Springer, New York, 2013.
[44] B. Hidalgo, M. Goodman, Multivariate or multivariable regression? Am. J. Public. Health. 103 (1) (2013) 39–40.
[45] W. Härdle, L. Simar, Applied Multivariate Statistical Analysis, 22007, Springer, Berlin, 2007.
[46] U. Grömping, Relative importance for linear regression in R: the package relaimpo, J. Stat. Softw. 17 (1) (2006) 1–27.
[47] R. Polikar, Ensemble learning, in: Ensemble Machine Learning, Springer US, 2012, pp. 1–34.
[48] S. Van Poucke, Z. Zhang, M. Schmitz, M. Vukicevic, V. Laenen, C. M., A. L., C. De Deyne, Scalable predictive analysis in critically ill patients using a visual open data analysis platform, PLoS One 11 (1) (2016) e0145791.
[49] Y. Freund, R. Schapire, N. Abe, A short introduction to boosting, J. -Jpn. Soc. Artif. Intell. 14 (771-780) (1999) 1612.
[50] V. Kempen, H. Kruize, H.C. Boshuizen, C.B. Ameling, B.A.M. Staatsen, A.E.M. de Hollander, The association between noise exposure and blood pressure and ischemic heart disease: a meta-analysis, Environ. Health. Perspect. 110 (3) (2002) 307–317.
[51] Z. Marton, F. Seidel, F. Balint-Benczedi, M. Beetz, Ensembles of strong learners for multi-cue classification, Pattern Recogn. Lett. 34 (May (7)) (2013) 754–761. http://dx.doi.org/10.1016/j.patrec.2012.07.011.
[52] A. Chamberlain, M. Paxton, K. Glover, M. Flintham, D. Price, C. Greenhalgh, S. Benford, P. Tolmie, E. Kanjo, A. Gower, A. Gower, D. Woodgate, Danaë Emma Beckford Stanton Fraser: understanding mass participatory pervasive computing systems for environmental campaigns, Pers. Ubiquitous Comput. 18 (7) (2014) 1775–1792.
[53] N. Alajmi, E. Kanjo, N.E. Mawass, Alan Chamberlain: Shopmobia: an emotion-based shop rating system, ACII (2013) 745–750.
[54] E. Kanjo, L. Al-barrak, E.M.G. Younis, NeuroPlace: categorizing urban places according to mental states, PLoS One (2017) in press.
[55] T. Münzel, T. Gori, W. Babisch, M. Basner, Cardiovascular effects of environmental noise exposure, Eur. Heart. J. 35 (13) (2014) 829–836, doi:10.1093/eurheartj/ehu030.
[56] U. Kraus, A. Schneider, S. Breitner, R. Hampel, R. Rückerl, M. Pitz, U. Geruschkat, P. Belcredi, K. Radon, A. Peters, Individual day-time noise exposure during routine activities and heart rate variability in adults: a repeated measures study, Environ. Health. Perspect. (2013). http://dx.doi.org/10.1289/ehp.1205606.
[57] N.J. Verberkmoes, M.A. Soliman Hamad, J.F. ter Woorst, T. MESH, C.H. Peels, A.H.M. van Straten, Impact of temperature and atmospheric pressure on the incidence of major acute cardiovascular events, Netherlands Heart Journal 20 (5) (2012) 193–196, doi:10.1007/s12471-012-0258-x.
[58] S. Danet, F. Richard, M. Montaye, S. Beauchant, B. Lemaire, C. Graux, D. Cottel, N. Marécaux, P. Amouyel, “Unhealthy Effects of Atmospheric Temperature and Pressure on the Occurrence of Myocardial Infarction and Coronary Deaths”, https://doi.org/10.1161/01.CIR.100.1.e1.
[59] E. Kanjo, L. Al-Husain, A. Chamberlain, Emotions in context: examining pervasive affective sensing systems, applications, and analyses, Pers. Ubiquitous Comput. 19 (7) (2015) 1197–1212.
[60] L. Al-Barrak, E. Kanjo, NeuroPlace: making sense of a place, AH (2013) 186–189.
[61] N.E. Mawass, E. Kanjo, in: A Supermarket Stress Map, UbiComp (Adjunct Publication), 2013, pp. 1043–1046.
[62] L. Al-Barrak, E. Kanjo, NeuroPlace: making sense of a place, AH (2013) 186–189.
[63] L. Al-Husain, E. Kanjo, A. Chamberlain, in: Sense of Space: Mapping Physiological Emotion Response in Urban Space, UbiComp (Adjunct Publication), 2013, pp. 1321–1324.
[64] Gno. Gravina, Automatic methods for the detection of accelerative cardiac defense response, IEEE Trans. Affective Comput. 7 (3) (2016) 286–298.
[65] M. Chen, S. González-Valenzuela, A.V. Vasilakos, H. Cao, V.C.M. Leung, Body area networks: a survey, MONET 16 (2) (2011) 171–193.
[66] G. Fortino, R. Giannantonio, R. Gravina, P. Kuryloski, R. Jafari, Enabling effective programming and flexible management of efficient body sensor network applications, IEEE Trans. Hum.-Mach. Syst. 43 (1) (2013) 115–133.
[67] B. Makivic, P. Bauer, Heart rate variability analysis in sport, Aspetar. Sports Med. J. 4 (September (2)) (2015) 326–331.